Appendix B. Finding Heavy Atoms or Anomalous Scatterers

This section demonstrates the use of the programs XPREP and XM (SHELXD) to determine the locations of the three insulin disulfide groups from the set of anomalous scattering data.

The sample data accompanying this manual is from a crystal of zinc-free bovine insulin. Zinc-free bovine insulin was chosen because:

- it is inexpensive and readily available
- crystal growth is easy and straightforward
- the crystal system is cubic which provides high redundancy in data measurements
- there are three disulfide (S-S) bonds in the structure

The goal of this exercise is to determine the positions of the three disulfide bonds from the single wavelength anomalous scattering (SAS) data produced by PROTEUM/ProScale. Sulfur has a weak (Δf" 0.56) anomalous scattering signal for SAS data recorded using Cu Ka radiation. However, by collecting highly redundant data, the anomalous signal-to-noise level in the resulting data set is increased to the point where it can be recorded accurately enough to yield both the positions of the anomalous scatterers and, by solvent flattening/noise filtering (Wang, 1985), the protein phases.

The 24-fold symmetry of the cubic zinc-free insulin crystals provides the needed data redundancy for the SAS analysis without needing to collect a large amount of data. Also, by using low (3.5 Å) data, the two sulfur positions in the disulfide bond appear as a single sulfur superatom which also aids in the analysis. This section briefly describes the procedures used for the analysis.

B.1 XPREP Setup

XPREP, written by George Sheldrick, is a versatile program for analyzing the scaled structure factors, merging data, producing Patterson maps, and preparing files for SHELXTL and XM.

B.1.1 Setup the Program on a Windows 2000/XP Computer

1. Copy the xprep.exe executable to the working directory.
2. To execute the program, double-click on the xprep.exe icon or use the Command Prompt to cd (change directory) to the working directory and then type xprep. The XPREP window will appear (see Figure B.1).

B.1.2 XPREP Window Prompts

1. Enter insulin_example_01_s.sca for the filename prompt.
2. Enter D for the Option prompt.
3. Enter Y for the Use from Denzo/Scalepack prompt.
4. Enter I for the Select option prompt. This will bring up the XPREP main window (see Figure B.2).

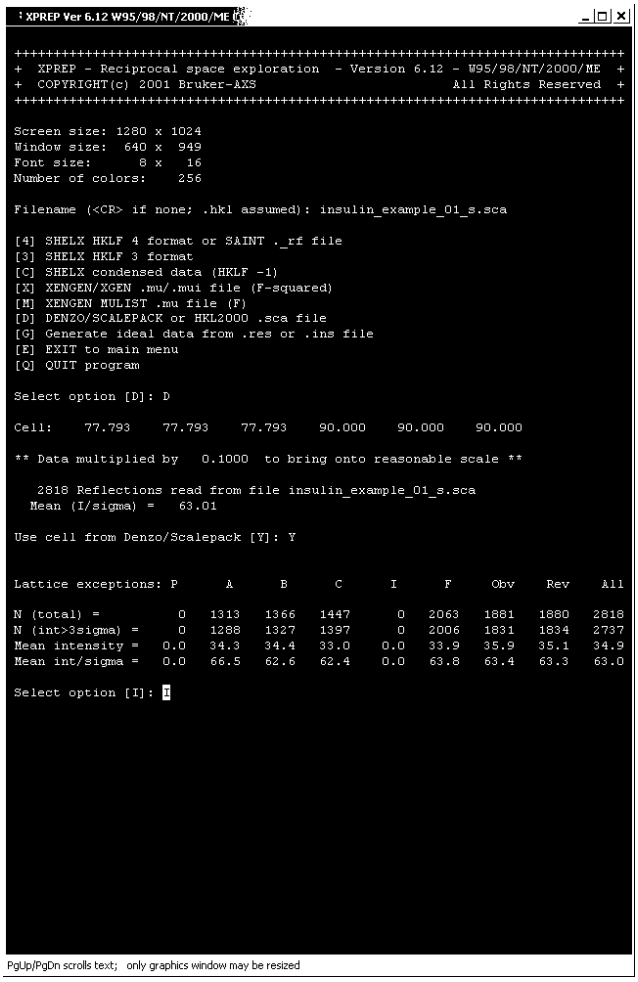
XPREP Ver 6.12 W95/98/NT/2000/ME
+ XPREP - Reciprocal space exploration - Version 6.12 - W95/98/NT/2000/ME +
+ COPYRIGHT(c) 2001 Bruker-AXS All Rights Reserved +
Screen size: 1280 x 1024
Window size: 640 x 949
Font size: 8 x 16
Number of colors: 256
Filename (<CR> if none; .hkl assumed): insulin_example_01_s.sca
[4] SHELX HKLF 4 format or SAINT ._rf file
[3] SHELX HKLF 3 format
[C] SHELX condensed data (HKLF -1)
[X] XENGEN/XGEN .mu/.mui file (F-squared)
[M] XENGEN MULIST .mu file (F)
[D] DENZO/SCALEPACK or HKL2000 .sca file
[G] Generate ideal data from .res or .ins file
[E] EXIT to main menu
[Q] QUIT program
Select option [D]: D
Cell: 77.793 77.793 77.793 90.000 90.000 90.000
** Data multiplied by 0.1000 to bring onto reasonable scale **
2818 Reflections read from file insulin_example_01_s.sca
Mean (I/sigma) = 63.01
Use cell from Denzo/Scalepack [Y]: Y
Lattice exceptions: P A B C I F Obv Rev All
N (total) = 0 1313 1366 1447 0 2063 1881 1880 2818
N (int>3sigma) = 0 1288 1327 1397 0 2006 1831 1834 2737
Mean intensity = 0.0 34.3 34.4 33.0 0.0 33.9 35.9 35.1 34.9
Mean int/sigma = 0.0 66.5 62.6 62.4 0.0 63.8 63.4 63.3 63.0
Select option [I]: I
PgUp/PgDn scrolls text; only graphics window may be resized

Figure B.1 - Utilities > XPREP

B.1.3 XPREP Main Window Prompts

1. Enter S for the Select option [S] prompt to determine the space group. This will refresh the XPREP main window.
2. Enter C for the Select option [C] prompt.
3. Enter I for the Select option [I] prompt.
4. Enter C for the Select option [B] prompt to define the space group as I 21 3 and to refresh the XPREP main window.
5. In the refreshed XPREP main window, enter D to the Select option [H] prompt. This will pop-up the Read, Modify or Merge window (see Figure B.3).

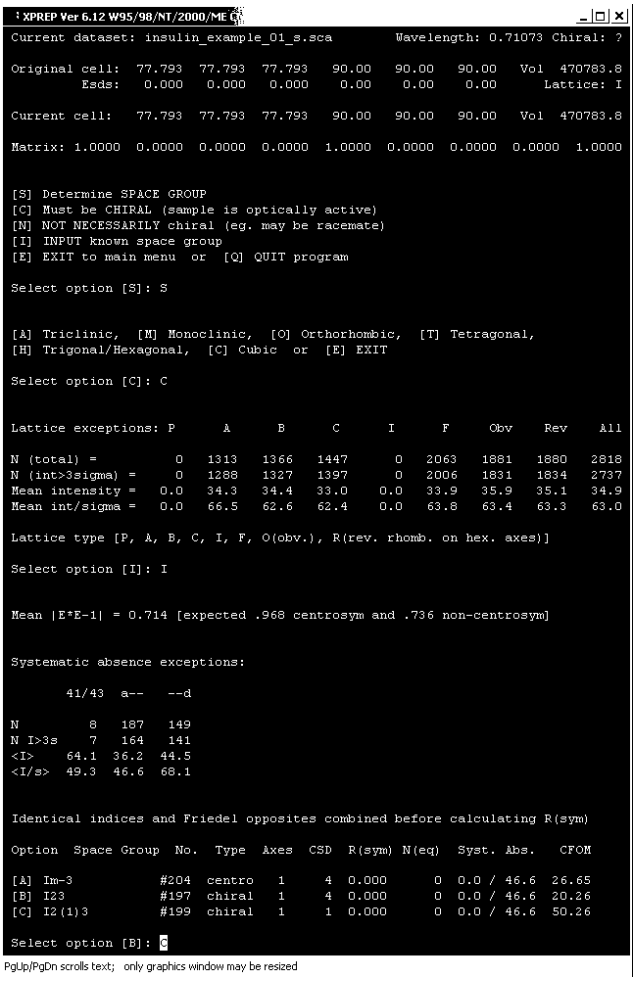
XPREP Ver 6.12 W95/98/NT/2000/ME
Current dataset: insulin_example_01_s.sca Wavelength: 0.71073 Chiral: ?
Original cell: 77.793 77.793 77.793 90.00 90.00 90.00 Vol 470783.8
Esds: 0.000 0.000 0.000 0.00 0.00 0.00 Lattice: I
Current cell: 77.793 77.793 77.793 90.00 90.00 90.00 Vol 470783.8
Matrix: 1.0000 0.0000 0.0000 0.0000 1.0000 0.0000 0.0000 0.0000 1.0000
[S] Determine SPACE GROUP
[C] Must be CHIRAL (sample is optically active)
[N] NOT NECESSARILY chiral (eg. may be racemate)
[I] INPUT known space group
[E] EXIT to main menu or [Q] QUIT program
Select option [S]: S
[A] Triclinic, [M] Monoclinic, [O] Orthorhombic, [T] Tetragonal,
[H] Trigonal/Hexagonal, [C] Cubic or [E] EXIT
Select option [C]: C
Lattice exceptions: P A B C I F Obv Rev All
N (total) = 0 1313 1366 1447 0 2063 1881 1880 2818
N (int>3sigma) = 0 1288 1327 1397 0 2006 1831 1834 2737
Mean intensity = 0.0 34.3 34.4 33.0 0.0 33.9 35.9 35.1 34.9
Mean int/sigma = 0.0 66.5 62.6 62.4 0.0 63.8 63.4 63.3 63.0
Lattice type [P, A, B, C, I, F, O(obv.), R(rev. rhomb. on hex. axes)]
Select option [I]: I
Mean |E*E-1| = 0.714 [expected .968 centrosym and .736 non-centrosym]
Systematic absence exceptions:
41/43 a-- --d
N 8 187 149
N I>3s 7 164 141
<I> 64.1 36.2 44.5
<I/s> 49.3 46.6 68.1
Identical indices and Friedel opposites combined before calculating R(sym)
Option Space Group No. Type Axes CSD R(sym) N(eq) Syst. Abs. CFOM
[A] Im-3 #204 centro 1 4 0.000 0 0.0 / 46.6 26.65
[B] I23 #197 chiral 1 4 0.000 0 0.0 / 46.6 20.26
[C] I2(1)3 #199 chiral 1 1 0.000 0 0.0 / 46.6 50.26
Select option [B]: C
PgUp/PgDn scrolls text; only graphics window may be resized

Figure B.2 - XPREP main window

B.1.4 Read, Modify or Merge Window Prompts

1. Enter A to the Select option [S] prompt. This will refresh the window.
2. Enter A to the Select option [E] prompt. Use the default parameters (by hitting <CR>) when prompted for the high resolution cut-off, I/_ cutoff, and target number of reflections.

XPREP will then display a table of the anomalous signal-to-noise ratios as a function of resolution. This table is useful in determining the resolution cutoff for the anomalous scattering data (i.e., the point at which the signal-to-noise ratio becomes random, in this case 3.8 Å).

3. Continue by using the defaults for the high resolution limit for output data and effective B-value.
4. At the Filename prompt, enter insulin_example.
5. Enter Y to the Write .ins file prompt.
6. Accept the filename by clicking <CR> at the Filename [insulin_example.ins]: prompt.
7. Enter S at the Element type for heavy atoms [Se] prompt and 2/3 at the Number of unique heavy atoms [12] prompt.

XPREP Ver 6.12 W95/98/NT/2000/ME
Index # Data Filename or Source of Data
1 5249 insulin_example_01_s.sca <- current dataset
[M] Sort-MERGE current data (no scaling)
[C] Change CURRENT dataset
[L] LEAST-SQUARES scale and merge datasets
[W] WRITE dataset to file
[I] INCLUDE Rfree flags from another file
[R] READ in another dataset
[S] Display intensity STATISTICS
[D] DELETE stored dataset
[F] FACE-indexed absorption corrections
[P] PSI-scan absorption corr.
[T] Copy file, TRANSFORM hkl and cosines
[A] MAD, SAS, SIR or SIRAS
[H] Apply HIGH/low resolution cutoffs
[N] NORMALIZE/scale sigmas
[G] Generate simulated powder diagrams
[E] EXIT to main menu
[Q] QUIT program
Select option [S]: A
[M] MAD (Multiple-wavelength Anomalous Dispersion)
[I] SIR (Single Isomorphous Replacement)
[A] SAS (Single-wavelength Anomalous Scattering)
[R] SIRAS (Single Isomorphous Replacement with Anomalous Scattering)
[E] EXIT to main menu
[Q] QUIT program
Select option [E]: A
High resolution limit in Angstroms for this calculation [0.0]: 0.0
I/sigma threshold for rejecting (after merging) [0.5]: 0.5
Target number of reflections in local scaling sphere
(0 if no local scaling) [100]: 100
Anomalous signal/noise ratios (1.0 is random). The first line is based on
input sigmas, the second on variances of F+ and F- (if not already averaged):
Inf - 8.0 - 6.0 - 5.0 - 4.1 - 3.9 - 3.7 - 3.5 - 3.3 - 3.1 - 2.9 - 2.7 - 2.5 A
3.21 3.49 2.69 2.37 1.77 1.86 1.58 1.64 1.55 1.37 1.13 1.15
66.4 Neighbors used on average for F+/F- local scaling
Rint(anom) = 0.0303 before and 0.0296 after local scaling
High resolution limit in Angstroms for saved data [0.000]: 0.0
Enter effective B-value (e.g. 20) to normalize delta-F or Fa values,
<CR> for no renormalization:
Filename to write FA and phi(T)-phi(A) (<CR> for none): INSULIN_EXAMPLE
Current dataset contains 2417 SAS delta(F)
Write .ins file for SHELXD/XD (Y or N)? [Y]: Y
Filename [INSULIN_EXAMPLE.ins]:
Element type for heavy atoms [Se]: S
Number of unique heavy atoms [12]: 3
Wavelength [1.54178]:
PgUp/PgDn scrolls text; only graphics window may be resized

Figure B.3 - XPREP read, modify, or merge window

8. Accept the wavelength by clicking <CR> at the Wavelength [1.54178] prompt. The contents of the insulin_example.ins file will display.
9. Enter <CR> to exit the window, which will refresh the XPREP Read Options window.
10. Enter E at the Option [S] prompt to return to the main window and to view the Patterson sections.

    Otherwise, enter Q at the Option [S] prompt to exit the program.
11. To display the anomalous difference Patterson maps, enter P at the Select option [H] prompt to access the Patterson window (see Figure B.4).

B.1.5 Patterson Window Options

1. Enter R at the Select option [E] prompt to set the resolution limits.
2. Enter 99. and 3.5 at the prompt to set the low and high resolution limits.
3. Enter O at the Select option [E] prompt to remove the origin.
4. Enter Z at the Select option [E] prompt to display the Patterson Z-section.
5. Enter 0.0 at the Z-coordinate prompt to display the Z = 0 Patterson section.
6. Enter <CR> to use the default for the Z-coordinate for sigma estimation. A list of Patterson peaks on the Z = 0 section display. Use Page Up and Page Down to scroll through the peak list.
7. Enter <CR> to display the contoured Patterson section (see Figure B.5).

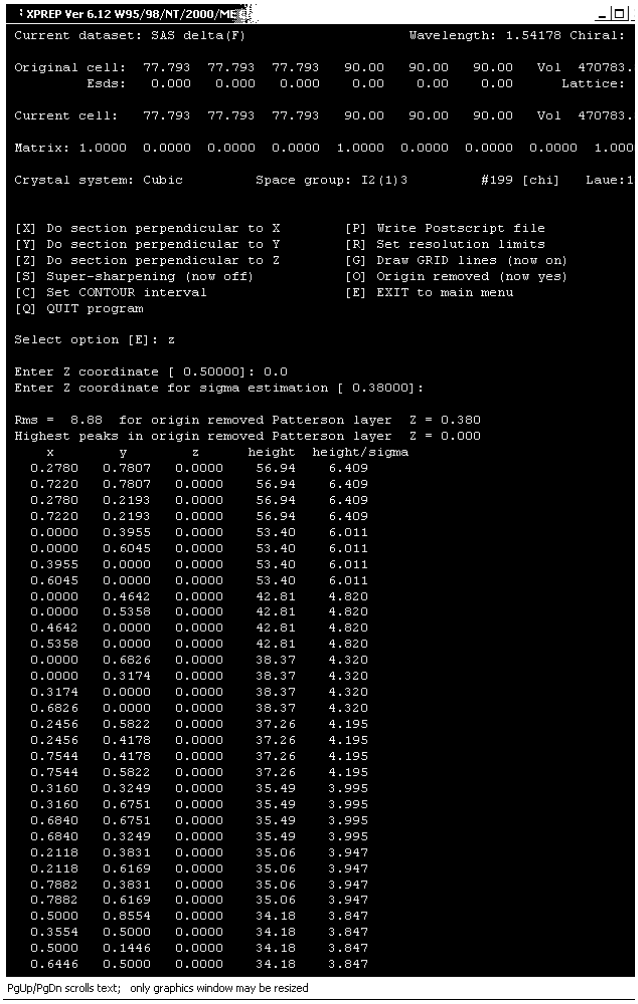

Figure B.4 - XPREP Patterson window

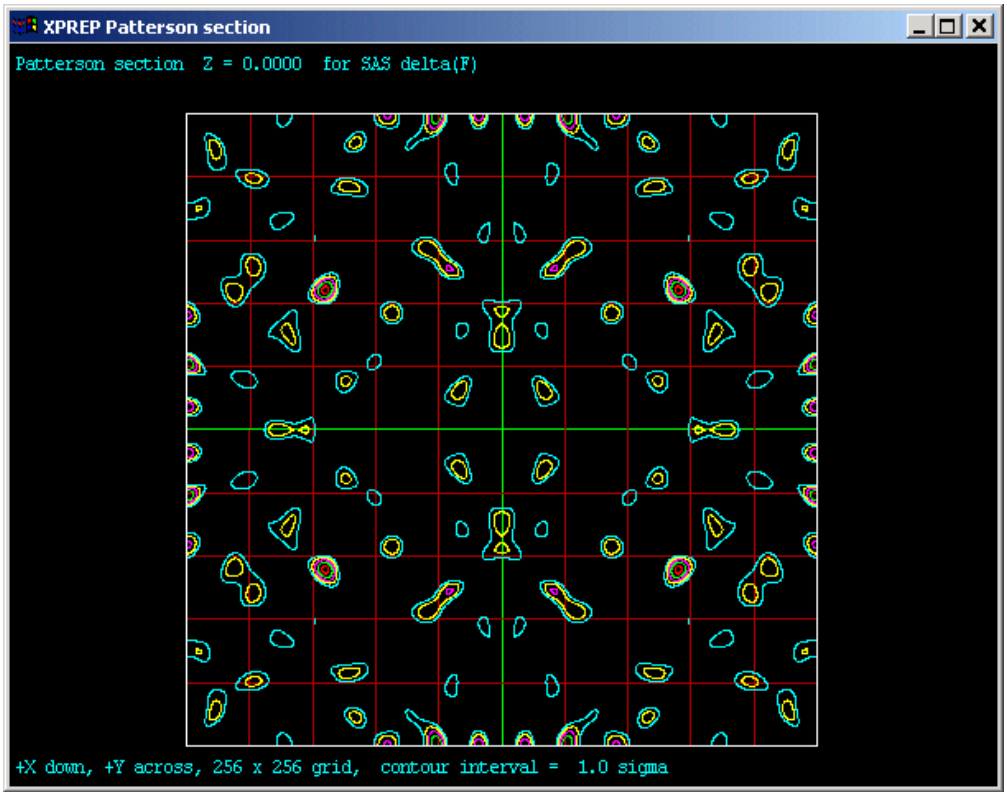
XPREP Patterson section
Patterson section Z = 0.0000 for SAS delta(F)
+X down, +Y across, 256 x 256 grid, contour interval = 1.0 sigma

Figure B.5 - Contoured Patterson section

The Patterson section for the insulin sulfur-SAS example data is noisy due to the weak sulfur signal and the limited redundancy of the 90° data set. However, the data is sufficient to locate the three disulfide positions using XM (SHELXD).

B.2 XM (SHELXD)

XM (SHELXD), written by George Sheldrick, is an extremely efficient program for locating heavy atoms or anomalous scatterers using the anomalous difference (*.hkl) and input (*.ins) files generated by XPREP.

For the insulin_example, you must modify the XM input file by adding three additional parameters (see Figure B.6).

```
TITL INSULIN_EXAMPLE in I2(1)3
CELL 1.54178  77.7930  77.7930  77.7930  90.000  90.000  90.000
ZERR   48.00   0.0090   0.0090   0.0090   0.000   0.000   0.000
LATT -2
SYMM 0.5-X, -Y, 0.5+Z
SYMM -X, 0.5+Y, 0.5-Z
SYMM 0.5+X, 0.5-Y, -Z
SYMM Z, X, Y
SYMM 0.5+Z, 0.5-X, -Y
SYMM 0.5-Z, -X, 0.5+Y
SYMM -Z, 0.5+X, 0.5-Y
SYMM Y, Z, X
SYMM -Y, 0.5+Z, 0.5-X
SYMM 0.5+Y, 0.5-Z, -X
SYMM 0.5-Y, -Z, 0.5+X
SFAC S
UNIT   144
PATS
FIND   3
MIND -3.5
SHEL 99. 3.5
NTRY 20
WEED 0
HKLF 3

END
```

Figure B.6 - XM input file

1. Insert the SHEL 99. 3.5 command line to modify the resolution range of the analysis to use data in the range of 99. to 3.5 Å resolution.
2. Insert the NTRY 20 command line to limit the total number of solutions to twenty. Omitting this line will allow the program to run indefinitely.
3. Insert the WEED 0 command line to speed up the calculations, since we are looking for only three sulfur super atoms.

NOTE: Since the expected XM input file has the .ins extension, Windows interprets this as a special file. Use the Open with the option to edit the file.

B.2.1 Start XM

1. Copy xm.exe to the working directory and run the program from there.
2. Type the XM filename (no extension) in the command prompt window to execute the program. The program will display output in the Command Prompt window, as well as write a copy of the output to the disk as file-name.lst (see Figure B.7).

```
PSUM 57.86  PSMF Peaks: 63 61 43 17 13 10 4 3
 Try 9:20  Peaks 99 96 94 35 34
 R = 0.334, Min.fun. = 0.472, <cos> = 0.341, Ra = 0.339
 Try     9, CC All/Weak 36.91 / 18.36, best 36.91 / 18.36, best
PATFOM 40.51

 PATFOM  40.51

    x       y       z     sof  U   height
 0.48441 0.33655 0.23803 20.0  1.0  99.90
 0.64152 0.63899 0.07690 20.0  1.0  98.43
 0.46549 0.91776 0.16196 20.0  1.0  97.60
 0.78958 0.18511 0.04419 20.0  1.0  59.45
 0.22447 0.02254 0.12924 20.0  1.0  58.95

 Minimum distances (top row, 0 if special position) and PSMF (bottom
row)

 Peak    x      y      z      self  cross-vectors

 99.9  0.4844 0.3365 0.2380   13.7
                              18.4

 98.4  0.6415 0.6390 0.0769   20.7   9.2
                              70.2  47.4

 97.6  0.4655 0.9178 0.1620   18.7  10.8 11.5
                              65.5  29.0 12.5

 -------------------------------------------------

 59.4  0.7896 0.1851 0.0442    9.2  18.3 18.0 15.6
                               0.0   0.0  0.0  0.0

 59.0  0.2245 0.0225 0.1292   19.1  23.7 16.8 20.6  9.6
                              19.1   0.0  0.0  0.3 12.7  0.0
```

inter-peak distance

Figure B.7 - Section of the XM output file

Since XM does not try to identify whether a solution has been found, there can be several correct solutions among the twenty requested. The best solution is the last solution where the Patterson peak list, self- and cross-vector peak heights, and inter-peak distances are output. The all-data and weak-data correlation coefficients (CC) are a sensitive indicator of a probable solution. In general, the solution is valid if the correlation coefficient for all-data is greater than twenty and the correlation for weak-data is greater than ten. Correct solutions should correspond to those with the largest Patterson figure of merit PATFOM. Also, there should be a good separation of peak heights in the list between the expected peaks and the noise. The Patterson self- and cross-vectors should be large and the inter-peak distance should be reasonable.

Zinc-free insulin belongs to space group I 2_1 3. In this space group, two non-symmetric, related, but equivalent, indexings are possible, resulting in forty-eight coordinate points for each of the three sulfur super atom locations (XM may also give an inverted xyz solution). This may lead to confusion if you want to compare the XM solutions to the zinc-free insulin PDB entry (9INS). The disulfide geometry or inter-peak distances presented in Figure B.7 correspond to the correct insulin disulfide geometry irrespective of indexing. Use this to check whether the XM solution from your own data is correct.